INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal, Hyderabad - 500 043

COMPUTER SCIENCE AND ENGINEERING (AI & ML)

QUESTION BANK

Course Title	**PREDICTIVE DATA ANALYTICS**				
Course Code	ACAC17				
Program	B.Tech				
Semester	VII	CSE(AI & ML)			
Course Type	Elective				
Regulation	IARE - UG 20				
Course Structure	Theory			Practical	
	Lecture	Tutorials	Credits	Laboratory	Credits
	3	-	3	-	-
Course Coordinator	Ms. S.Viharika, Assistant Professor				

COURSE OBJECTIVES:

The students will try to learn:

I	The fundamental concepts of data analytics. .
II	The principles and methods of statistical analysis. .
III	The models used in predictive data analytics applications using supervised machine learning.
IV	The interesting patterns analyze supervised and unsupervised models and estimate the accuracy of the algorithms.

COURSE OUTCOMES:

After successful completion of the course, students should be able to:

CO 1	**Demonstrate** the fundamental concepts and types of machine learning its role in predictive data analytics and real-world applications	Understand
CO 2	**Make use of** descriptive and exploratory analysis methods to extract valuable insights from data to provide a foundation for decision-making.	Apply
CO 3	**Build** a techniques to collect relevant data and assess its quality tensuring the foundation for accurate information extraction	Apply

CO 4	**Illustrate** probability-based learning to develop probabilistic classification and regression models capturing uncertainty in predictions..	Understand
CO 5	**Utilize** the principles of error-based learning and its significance in improving models through the analysis of prediction errors.	Apply
CO 6	**Identify** the foundational concepts of machine learning, including types of algorithms, training, and prediction and their application in predictive data analytics.	Apply

QUESTION BANK:

Q.No	QUESTION	Taxonomy	How does this subsume the level	CO's
MODULE I				
MACHINE LEARNING FOR PREDICTIVE DATA ANALYTICS				
PART A-PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS				
1	Are there ethical concerns related to the use of predictive analytics, especially regarding privacy or discrimination? What legal regulations or compliance requirements should be adhered to in the data analytics process? How can ethical and legal issues be addressed while still achieving project objectives?	Apply	The learner will try to **recall** privacy and bias concerns, **Understand** why bias in predictive models is problematic and how it can impact different groups and then **Apply** ethical principles in the design and implementation of predictive analytics projects to ensure compliance and fairness.	CO 1
2	Do the tools provide ways to interpret and explain the predictions made by the models? How can we ensure that the models' predictions are transparent and understandable to stakeholders?	Analyze	The learner will try to **recall** the selection, implementation, and utilization of **Understand** predictive data analytics tools,to **Apply** and **Analyze** ultimately leading to more successful and impactful data-driven projects.	CO 1

3	Are there alternative predictive data analytics tools or platforms that should be considered? What are the trade-offs and advantages of each alternative?	Understand	The learner will try to **recall** the selection, implementation, and utilization, **Understand** predictive data analytics tools, ultimately leading to more successful and impactful data-driven projects.	CO 1
4	What ethical considerations arise from the data, insights, and decisions?Are there social, legal, or cultural implications that need to be addressed? How can ethical principles be integrated into the decision-making process?	Apply	The learner will try to **recall** data to insights decisions, **Understand** organizations can make more informed and effective choices and then **Apply** the power of data analytics to drive positive outcomes.	CO 1
5	Is the technology or methodology needed for the project available and proven?Are there technical risks or challenges that need to be addressed?Do we have the necessary expertise to execute the project from a technical standpoint?	Understand	The learner will try to **recall** the feasibility of a project comprehensively, **Understand** enabling organizations to make informed decisions and mitigate risks before committing resources and efforts to the initiative.	CO 1
6	Is there a market or demand for the project's outcomes or deliverables?Have market trends and customer needs been analyzed to support the project's viability?Are there competitors or similar initiatives in the market, and how does our project compare?	Analyze	The learner will try to **recall** the feasibility of a project comprehensively, **Understand** enabling organizations to make informed decisions, **Apply** mitigate risks and **Analyze** before committing resources and efforts to the initiative.	CO 1

7	Are there legal or regulatory requirements that must be met to execute the project?How can compliance with relevant laws and regulations be ensured?Are there potential legal risks associated with the project?	Apply	The learner will try to **recall** the feasibility of a project comprehen-sively,**Understand** enabling organizations to make informed decisions and **Apply** mitigate risks before committing resources and efforts to the initiative.	CO 1
8	Are there alternative approaches or strategies that could achieve the same objective with lower risk or cost? What contingency plans are in place in case the project faces unexpected challenges or obstacles? How flexible is the project plan to adapt to changing circumstances?	Apply	The learner will try to **recall** the feasibility of a project comprehensively, **Understand** enabling organizations to make informed decisions and then **Apply** mitigate risks before committing resources and efforts to the initiative.	CO 1
9	Do categorical variables need to be encoded (e.g., one-hot encoding)?Should numerical variables be scaled or normalized to ensure consistent model performance?Are there text or unstructured data that require specialized processing (e.g., NLP techniques)?	Analyze	The learner will try to **recall** the concepts, **understand** why these preprocessing steps are necessary, **Apply** them based on your data and problem, and **Analyze** the results to determine if the chosen techniques are improving your model's performance.	CO 1
10	Are there missing values or inconsistent entries in customer records that need cleaning? How can outliers in usage data be identified and handled appropriately?	Apply	The learner will try to **recall**data cleaning and handling outliers are crucial steps in data preprocessing, **Understand** their significance and **Apply**apply the appropriate techniques to ensure that your data is of high quality and suitable for analysis and modeling.	CO 1
PART-B LONG ANSWER QUESTIONS				

1	What is algorithmic bias, and how can it affect the outcomes of predictive analytics models?Explain the concept of fairness in machine learning and its ethical implications.Can you discuss strategies and approaches for detecting and mitigating bias in predictive models?	Understand	The learner will try to **recall** involves remembering the concepts and terminology related to algorithmic bias and **understand** entails grasping why these issues are important and the potential consequences of ignoring them.	CO 1
2	How do organizations transition from model development to deploying predictive analytics solutions in a production environment? Provide case studies or examples of real-world applications of predictive data analytics in various industries, such as healthcare, finance, and e-commerce. What are the challenges and considerations when scaling predictive analytics solutions for large-scale deployment?	Apply	The learner will try to **recall** involves remembering key concepts and steps in the transition from model development to production deployment,**Understand** the importance of this transition and the challenges it entails and then **Apply** using appropriate strategies and best practices to successfully deploy and scale predictive analytics solutions in real-world settings.	CO 1
3	Can you explain the significance of the Business Understanding phase in the CRISP-DM framework?How does aligning the data analytics project with the business objectives contribute to project success?	Understand	The learner will try to **recall** the significance of the Business Understanding phase in CRISP-DM and then **Understand** the importance of aligning data analytics projects with business objectives.	CO 1

4	What is the primary purpose of the Data Understanding phase, and how does it inform subsequent stages?Describe the key activities involved in data collection, data description, and data exploration.How can data profiling and data quality assessment help identify potential issues in the dataset?	Analyze	The learner will try to **recall** the importance of the Data Understanding phase and its role in shaping the success of a data analytics project, **understand** the purpose and key activities involved in this phase, **Apply** entails using data profiling and quality assessment techniques to identify and address potential data issues and then**Analyze** encompasses the process of gaining insights into the dataset through exploratory data analysis, which informs subsequent stages of the project.	CO 1
5	How does the CRISP-DM framework address the need for continuous improvement in predictive data analytics projects?Explain the concept of model drift and how it can impact the performance of deployed models.	Understand	The learner will try to **recall** CRISP-DM promotes continuous improvement in predictive data analytics and **Understand** addressing model drift is a critical aspect of maintaining the performance and relevance of deployed models over time.	CO 2

6	Discuss the role of predictive analytics tools in model evaluation and validation.How do these tools help users choose relevant evaluation metrics and visualize model performance?What features or techniques are available for improving the interpretability of complex predictive models?	Evaluate	The learner will try to **recall** the significance of using predictive analytics tools , **understand** how these tools assist in evaluating and validating predictive models, **Apply** these tools to effectively assess and refine models,**Analyze** the results generated by the tools to make informed decisions and then **evaluate** the overall model performance and effectiveness using these tools.	CO 1
7	Explain the process of model development within predictive data analytics tools.How do these tools facilitate the selection of appropriate machine learning algorithms for specific tasks?Can you describe the advantages and limitations of using automated machine learning (AutoML) tools for model building?	Understand	The learner will try to **recall** these strategies can help students gain a deep **understand** of the model development process, algorithm selection, and the benefits and limitations of AutoML tools in predictive data analytics.	CO 1
8	Discuss the role of predictive analytics tools in model evaluation and validation.How do these tools help users choose relevant evaluation metrics and visualize model performance?What features or techniques are available for improving the interpretability of complex predictive models?	Evaluate	The learner will try to **recall** these strategies can help students **understand** the significance **Apply** and **Analyze** predictive analytics tools in model **Evaluate** validation and develop practical skills for model assessment and interpretation.	CO 1

9	How are predictive data analytics tools leveraging artificial intelligence (AI) and automation to streamline the analytics process?Can you provide examples of AI-driven features that enhance predictive modeling and data analysis?What are the implications of increased automation in predictive analytics for data scientists and analysts?.	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the evolving landscape of predictive analytics and the role of AI-driven automation in transforming data analysis and modeling processes.	CO 1
10	Explain the importance of data preparation and cleaning in the data-to-insights-to-decisions process.What steps are typically involved in data cleaning, and how do they help in obtaining meaningful insights?Share examples of common data preprocessing tasks and their impact on decision-making.	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the significance of data preparation and cleaning in the data-to-insights-to-decisions process and equip them with practical skills for handling real-world data.	CO 1
11	Discuss the ethical considerations related to the data-to-insights-to-decisions process, including privacy and bias.How can organizations ensure ethical data practices when making decisions based on data insights?Can you provide examples of ethical dilemmas and how they were addressed in decision-making?	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the significance of data preparation and cleaning in the data-to-insights-to-decisions process and equip them with practical skills for handling real-world data.	CO 1

12	Explain the importance of data visualization in conveying insights to stakeholders.What principles should be followed to create effective data visualizations for decision-makers?Can you share examples of impactful data visualizations that influenced critical decisions?.	Understand	The learner will try to **recall**these strategies, educators can help students appreciate the significance **Understand** data visualization and develop the skills needed to create effective visualizations that influence critical decisions.	CO 1
13	How does exploratory data analysis contribute to understanding the underlying patterns and relationships in data? Describe the techniques and visualizations used in EDA to uncover insights.Can you provide real-world examples where EDA led to unexpected discoveries with significant implications for decision-making?	Understand	The learner will try to **recall**these strategies, educators can help students appreciate the significance **Understand** data visualization and develop the skills needed to create effective visualizations that influence critical decisions.	CO 1
14	How can organizations measure the long-term impact of decisions made based on data insights? - What strategies can be implemented to continuously learn from data and improve decision-making processes?	Understand	The learner will try to **recall**these strategies, educators can help students appreciate the significance **Understand** data visualization and develop the skills needed to create effective visualizations that influence critical decisions.	CO 1
15	What is the significance of aligning project objectives with organizational goals in the feasibility assessment process? Can you provide examples of how misalignment between project objectives and organizational objectives can lead to challenges?	Understand	The learner will try to **recall** the various facets of feasibility assessment **Understand** emphasizing its importance in informed decision-making for a wide range of projects and initiatives	CO 1

16	Explain the components of financial feasibility analysis, including cost estimation, revenue projections, and return on investment (ROI) calculations.How can organizations determine the financial viability of a project and ensure it aligns with budgetary constraints?	Understand	The learner will try to **recall** the various facets of feasibility assessment **Understand** emphasizing its importance in informed decision-making for a wide range of projects and initiatives	CO 1
17	How do legal and regulatory factors impact the feasibility of a project?What strategies should organizations employ to ensure compliance with relevant laws and regulations?	Understand	The learner will try to **recall**The learner will try to **recall** the various facets of feasibility assessment **Understand** emphasizing its importance in informed decision-making for a wide range of projects and initiatives.	CO 1
18	How does project duration and the proposed schedule influence feasibility assessment?Explain the methods for estimating project timelines and potential delays.Share experiences from projects where time constraints played a pivotal role in feasibility decisions.	Understand	The learner will try to **recall** the various facets of feasibility assessment **Understand** emphasizing its importance in informed decision-making for a wide range of projects and initiatives.	CO 1
19	Explain the primary purpose of creating an Analytics Base Table (ABT) in data analytics projects? How does the design of the ABT impact the overall success of the analytics project?Provide examples of situations where a well-designed ABT significantly improved the efficiency of data analysis.	Understand	The learner will try to **recall**comprehensive exploration of the ABT design process, emphasizing .**Demonstrate** its importance in data analytics projects and showcasing the practical applications through a case study.	CO 1

20	How does variable selection and feature engineering contribute to the effectiveness of an ABT?Describe the techniques and considerations for identifying relevant variables and creating new features.Provide case studies where feature engineering led to better predictive models or more insightful analysis.	Understand	The learner will try to **recall**comprehensive exploration of the ABT design process, emphasizing .**Demonstrate** its importance in data analytics projects and showcasing the practical applications through a case study.	CO 1
PART-C SHORT ANSWER QUESTIONS				
1	What is the difference between supervised and unsupervised learning in Predictive Analytics?.	Understand	The learner will try to **recall** the choice between these two approaches depends on **Understand** the specific goals and characteristics of the data	CO 1
2	How can organizations benefit from Predictive Data Analytics in terms of decision-making?	Understand	The learner will try to **recall** by summarizing the key benefits of predictive data analytics **Understand** in decision-making	CO 1
3	Can the CRISP-DM framework be adapted to different industries and domains?	Understand	The learner will try to **recall** the CRISP-DM framework is a flexible and Understand widely used methodology for data mining and analytics projects.	CO 1
4	What are some advantages of using the CRISP-DM framework for predictive data analytics projects?	Understand	The learner will try to **recall** the CRISP-DM framework is a flexible and **Understand** widely used methodology for data mining and analytics projects	CO 1

5	How many main phases are there in the CRISP-DM framework?	Understand	The learner will try to **recall** the CRISP-DM framework is a flexible and widely used methodology for data mining and analytics projects, and **undersatnd** its adaptability makes it suitable for various applications.	CO 1
6	What is the first phase of the CRISP-DM framework, and what does it entail?	Understand	The learner will try to **recall** the CRISP-DM framework is a flexible and widely used methodology for data mining and analytics projects, and **undersatnd** its adaptability makes it suitable for various applications.	CO 1
7	Can you name a few widely used Predictive Data Analytics Tools?	Understand	The learner will try to **recall** these tools by summarizing the names and a brief description of each tool **understand** emphasizing their importance in the field of predictive data analytics.	CO 1
8	What types of tasks can Predictive Data Analytics Tools perform?	Understand	The learner will try to **recall** these tools by summarizing the names and a brief description of each tool **understand** emphasizing their importance in the field of predictive data analytics.	CO 1
9	How do organizations benefit from using Predictive Data Analytics Tools?	Understand	The learner will try to **recall** these tools by summarizing the names and a brief description of each tool **understand** emphasizing their importance in the field of predictive data analytics	CO 1

10	What is the difference between open-source and commercial Predictive Data Analytics Tools?	Understand	The learner will try to **recall** the choice between them depends on an organization's budget **understand** the level of support and customization needed.	CO 1
11	How do Predictive Data Analytics Tools handle data integration from various sources?	Understand	The learner will try to **recall** the choice between them depends on an organization's budget **understand** the level of support and customization needed.	CO 1
12	Explain the considerations when selecting the right Predictive Data Analytics Tool for a specific project?	Understand	The learner will try to **recall** these strategies into the curriculum, educators can prepare students to make informed decisions **Understand** when selecting predictive data analytics tools for their future projects.	CO 1
13	Define "Data Insights."	Understand	The learner will try to **recall** these strategies refer to valuable, actionable, and **understand**often previously undiscovered information from data analysis.	CO 1
14	How does data visualization contribute to the understanding of data and insights?	Understand	The learner will try to **recall** these strategies refer to valuable, actionable, and **understand**often previously undiscovered information from data analysis.	CO 1
15	What is the primary focus of exploratory data analysis (EDA) in the journey from data to insights?	Understand	The learner will try to **recall** the primary focus of EDA is to gain an initial **understand** of the dataset's characteristics, identify potential insights, and guide further data analysis.	CO 1

16	How do machine learning models contribute to generating predictive insights from data?	Understand	The learner will try to **recall** the primary focus of EDA is to gain an initial **understand** of the dataset's characteristics, identify potential insights, and guide further data analysis	CO 1
17	Explain the difference between descriptive and predictive insights.	Understand	The learner will try to **recall** Descriptive insights look at the past to explain what happened and why **understand** while predictive insights focus on the future, forecasting what will happen based on historical data and patterns.	CO 1
18	What is feasibility assessment in project management and decision-making?	Understand	The learner will try to **recall** the concept of feasibility assessment **understand** in project management and decision-making	CO 1
19	How does financial feasibility evaluation determine the project's viability?	Understand	The learner will try to **recall** the concept of feasibility assessment **understand** in project management and decision-making	CO 1
20	Explain the role of data selection in ABT design..	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the critical role of data selection in ABT design and prepare them to make informed decisions when conducting A/B tests in their future roles.	CO 1
MODULE II				
DATA TO INSIGHTS TO DECISION				
PART-A PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS				

1	In what situations is it crucial for models to be interpretable and explainable, and how can organizations achieve this? What techniques can be employed to make complex machine learning models more understandable to non-technical stakeholders?	Analyze	The learner will try to **recall** These critical thinking and problem-solving questions **Understand** address the multifaceted challenges **Apply** and **Analyze** considerations involved in the journey from data to insights to decisions.	CO 2
2	When is it appropriate to combine quantitative data with qualitative insights, and how can organizations effectively blend these sources of information? What challenges might arise when integrating qualitative data into a primarily quantitative analysis process?	Understand	The learner will try to **recall**These critical thinking and problem-solving questions **Understand** address the multifaceted challenges and considerations involved in the journey from data to insights to decisions.	CO 2
3	When dealing with categorical data, what are the various encoding methods available, and how do you choose the most suitable one? How can you address the challenge of high-cardinality categorical variables in feature engineering?	Apply	The learner will try to **recall**These critical thinking and problem-solving questions **Understand** address the complexity of feature engineering in **Apply** data science and machine learning.	CO 2
4	In what situations is dimensionality reduction through techniques like Principal Component Analysis (PCA) or t-SNE beneficial, and how do you decide to use them?What are the trade-offs between dimensionality reduction and retaining interpretability of features?	Apply	The learner will try to **recall**These critical thinking and problem-solving questions **Understand** address the complexity of feature engineering in **Apply** data science and machine learning.	CO 2

5	After implementing features, how can you monitor their performance and relevance over time, and when should feature engineering be revisited? What strategies can be employed to maintain and update feature engineering pipelines in production environments?	Understand	The learner will try to **recall**These critical thinking and problem-solving questions **Understand** address the complexity of feature engineering in data science and machine learning	CO 2
6	Consider situations where outliers or heavy-tailed distributions challenge the assumption of normality. What alternative distribution models or robust statistical methods can be employed?	Understand	The learner will try to **recall**critical thinking and problem-solving questions encourage a deeper **understand** of the normal distribution and its practical applications in real-world scenarios through a case study..	CO 2
7	Inconsistent data formats or coding schemes can lead to data quality problems. How can you standardize and clean data to address such issues?	Analyze	The learner will try to **recall**These critical thinking and problem-solving questions **Understand** encourage thoughtful consideration of data quality issues **Apply** and **Analyze** the development of strategies to address them effectively.	CO 2
8	Highlight the consequences of making decisions based on poor-quality data and how data quality improvements can lead to better decision-making.	Apply	The learner will try to **recall**These critical thinking and problem-solving questions **Understand** encourage thoughtful ensuring that data analysis **Apply** decision-making are based on reliable and high-quality data	CO 2

9	Consider when comparing data across different sources or time periods, what discrepancies or inconsistencies might indicate data quality problems?	Apply	The learner will try to **recall** problem-solving and critical thinking questions **Understand** provide a framework **Apply** for systematically identifying data quality issues	CO 2
10	In advanced data exploration, when and how can you use advanced imputation techniques like k-nearest neighbors imputation or interpolation to handle missing data??	Understand	The learner will try to **recall** problem-solving and critical thinking questions **Understand** encourage a deeper and more sophisticated approach to data exploration.	CO 2
PART-B LONG ANSWER QUESTIONS				
1	Explain the concept of EDA and its role in understanding data patterns, distributions, and potential outliers. Provide examples of visualizations and statistical techniques used in EDA?	Understand	The learner will try to **recall** long-answer questions **Understand** a comprehensive overview of the data-to-insights-to-decisions process.	CO 2
2	Outline the steps involved in selecting appropriate statistical models or machine learning algorithms for analysis. Discuss the evaluation and validation of models, including cross-validation and performance metrics.	Apply	The learner will try to **recall**emphasizing the key steps, **Understand**challenges, and **Apply** best practices involved in this critical journey.	CO 2
3	Explain the process of interpreting results, including identifying significant findings and assessing their practical implications. Discuss the challenges of interpreting complex models and high-dimensional data.	Understand	The learner will try to **recall** these strategies, educators can prepare students to tackle **Understand** the challenges of interpreting results, whether they involve complex models, high-dimensional data, or real-world decision-making scenarios.	CO 2

4	Discuss the ethical challenges that may arise during the data-to-insights-to-decisions process, such as bias, privacy, and transparency. Explain how organizations can address these challenges responsibly.	Understand	The learner will try to **recall** these strategies into the curriculum, educators can prepare students to navigate **Understand** the ethical challenges inherent in the data-to-insights-to-decisions process and make responsible and informed decisions in their future roles.	CO 2
5	Discuss the Role of Vision Transformers in Autonomous Systems and Robotics. How Can Transformers Enhance Visual Perception and Scene Understanding in Robots, Drones, and Autonomous Vehicles, Enabling Safe and Efficient Decision-Making?	Understand	The learner will try to **recall** long-answer questions provide a comprehensive overview of the data-to-insights-to-decisions process	CO 2
6	Explain how organizations can measure the impact and return on investment (ROI) of their data-driven decisions. Discuss the long-term benefits of data-driven decision-making for organizations.	Understand	The learner will try to **recall** long-answer questions **Understand** a comprehensive overview of the data-to-insights-to-decisions process	CO 2
7	Describe various encoding techniques for categorical features, including one-hot encoding, label encoding, and target encoding. What are their advantages and limitations?	Understand	The learner will try to **recall** long-answer questions **Understand** a comprehensive exploration of feature engineering, covering various aspects, techniques, and considerations in designing.	CO 2
8	Explain why the normal distribution is considered a fundamental and ubiquitous probability distribution in statistics.	Understand	The learner will try to **recall** the concept of then Information gain and Gini Index**Demonstrate** the implementation	CO 2

9	Discuss the moments of the normal distribution, including mean, variance, skewness, and kurtosis. How do these moments influence the shape and interpretation of the distribution?	Apply	The learner will try to **recall** these strategies, educators can help students grasp **understand**the importance of moments in describing and **Apply** interpreting probability distributions like the normal distribution and develop a deeper understanding of their role in statistical analysis.	CO 2
10	Discuss the importance of assessing the normality of data in the case study and methods for testing the assumption of normality.	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the significance of assessing the normality of data and equip them with the skills to perform these assessments using appropriate methods.	CO 2
11	How can statistical tests, such as the z-test or t-test, be applied in the case study to draw conclusions or make decisions based on the normal distribution?	Understand	The learner will try to **recall** these strategies, educators can help students grasp **understand**the importance of moments in describing and interpreting probability distributions like the normal distribution and develop a deeper understanding of their role in statistical analysis	CO 2
12	Explain the concept of data integrity and its significance in maintaining data quality. How does data integrity relate to data accuracy and completeness?	Understand	The learner will try to **recall** These long-answer questions **Understand** comprehensive exploration of data quality issues, their impact, assessment, and strategies for improvement.	CO 2

13	How can data quality dimensions such as accuracy, completeness, consistency, timeliness, and reliability be quantified and measured?	Remember	The learner will try to **recall** These long-answer questions provide a comprehensive exploration of data quality issues, their impact, assessment, and strategies for improvement.	CO 2
14	Explain the role of data validation rules and constraints in maintaining data quality. How can organizations establish and enforce data validation procedures?.	Understand	The learner will try to **recall** These long-answer questions **Understand** comprehensive exploration of data quality issues, their impact, assessment, and strategies for improvement.	CO 2
15	Discuss organizations ensure ongoing data quality and quality assurance, especially when dealing with large and constantly changing datasets?	Understand	The learner will try to **recall** These long-answer questions **Understand** comprehensive exploration of data quality issues, their impact, assessment, and strategies for improvement.	CO 2
16	Explain the concept of data as an organizational asset and how investing in data quality can yield long-term benefits and competitive advantages.	Understand	These long-answer questions **recall** these strategies, educators can help students grasp the concept of data as an organizational asset and **understand** how investing in data quality is a strategic decision with significant long-term benefits.	CO 2
17	Describe the steps involved in assessing data quality. How do data profiling, data auditing, and data validation contribute to this process?	Understand	These long-answer questions **recall** these strategies, educators can help students grasp the concept of data as an organizational asset and **understand** how investing in data quality is a strategic decision with significant long-term benefits	CO 2

18	Explain how these data quality challenges can lead to misinformation and incorrect decision-making. Provide real-world examples if possible.	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the real-world implications of data quality challenges and the importance of ensuring data accuracy, completeness, and reliability for informed decision-making.	CO 2
19	Define the concept of data governance and its relationship to data quality management. How can organizations establish effective data governance frameworks?	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the real-world implications of data quality challenges and the importance of ensuring data accuracy, completeness, and reliability for informed decision-making	CO 2
20	Explain the importance of data preparation in the context of data analysis and machine learning. Provide a step-by-step guide to the data preparation process, highlighting key considerations and techniques.	Understand	The learner will try to **recall** underlying mechanism, domain knowledge,**Understand** appropriate techniques to maintain data integrity and the validity of subsequent analyses.	CO 2
PART-C SHORT ANSWER QUESTIONS				
1	Why is data collection considered a crucial step in the process?	Understand	The learner will try to **recall** Data collection is considered a crucial step in **Undersatnd** the process because it forms the foundation for all subsequent stages of data analysis and decision-making.	CO 2

2	Give an example of feature engineering for a text-based analysis task.	Understand	The learner will try to **recall** these strategies into the curriculum, educators can help students grasp **Understand** the concept of feature engineering for text-based analysis tasks and equip them with practical skills.	CO 2
3	How does continuous improvement contribute to the organization's growth?	Remember	—	CO 2
4	What is polynomial feature engineering, and in what scenarios is it applied?	Understand	The learner will try to **recall** polynomial feature engineering is applied **Understand** when there are non-linear relationships between variables, and it can be useful in regression	CO 2
5	How can domain knowledge be valuable in feature engineering?	Understand	The learner will try to **recall** polynomial feature engineering is applied **Understand** when there are non-linear relationships between variables, and it can be useful in regression	CO 2
6	Explain the concept of feature selection in the context of machine learning.	Understand	The learner will try to **recall** these strategies into the curriculum, educators can help students grasp **Understand** the concept of feature engineering for text-based analysis tasks and equip them with practical skills for handling textual data effectively	CO 2
7	How can domain knowledge be valuable in feature engineering?	Remember	—	CO 2
8	What is the standard normal distribution, and why is it useful in statistical calculations?	Remember	—	CO 2

9	How is the z-score calculated, and what information does it provide about a data point in a normal distribution?	Remember	—	CO 2
10	In a case study involving exam scores, explain how the normal distribution can be applied to assess student performance.	Understand	The learner will try to **recall** the normal distribution can be applied to assess student performance by providing a statistical framework for **understand** how students are distributed across the score spectrum.	CO 2
11	How can outliers affect data quality, and what techniques are used to handle them?	Remember	—	CO 2
12	What is the significance of data profiling in assessing data quality?	Remember	—	CO 2
13	Explain the importance of data validation checks in data quality management.	Understand	The learner will try to **recall** these strategies data validation checks are essential in data quality management because they ensure data accuracy, integrity, compliance, and efficiency.	CO 2
14	How can data standardization contribute to data quality improvement?	Remember	—	CO 2
15	What is meant by "data governance," and how does it relate to data quality management?	Remember	—	CO 2
16	How can proactive data quality measures prevent data quality issues from occurring?	Remember	—	CO 2

17	Name a technique for visualizing high-dimensional data in advanced data exploration?	Understand	The learner will try to **recall** One technique for visualizing **understand** high-dimensional data in advanced data exploration is "t-SNE" or t-Distributed Stochastic Neighbor Embedding.	CO 2
18	In the context of advanced data exploration, what is feature selection, and why is it important?	Understand	The learner will try to **recall** One technique for visualizing **understand** high-dimensional data in advanced data exploration is "t-SNE" or t-Distributed Stochastic Neighbor Embedding.	CO 2
19	What role does outlier detection play in advanced data exploration, and why is it important?	Understand	The learner will try to **recall** its importance lies in identifying and **Understand** handling data points that deviate significantly from the typical patterns within a dataset.	CO 2
20	Explain the concept of feature engineering and its significance in advanced data exploration.	Understand	The learner will try to **recall** these strategies **Understand** Feature engineering is a crucial step in advanced data exploration and has significant importance.	CO 2
MODULE III				
INFORMATION BASED LEARNING				
PART A-PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS				

1	In the context of information-based learning, how does feature selection impact the quality of insights and models? Provide an example. What strategies can be employed to preprocess and clean data effectively to ensure that it's informative and free from biases?	Understand	The learner will try to **recall** critical thinking and problem-solving questions **Understand** explore various facets of information-based learning, from data acquisition to ethical considerations.	CO 3
2	When selecting a machine learning model for information-based learning, what factors should be considered, and how can one determine the most suitable model for a given problem?	Analyze	The learner will try to **recall**critical thinking and problem-solving questions **Understand** explore various facets of information-based learning **Apply** and **Analyze** from data acquisition to ethical considerations.	CO 3
3	In multiclass classification, how is Shannon entropy extended to measure the impurity or disorder in a set of classes? Discuss its role in algorithms like decision trees.	Understand	The learner will try to **recall**explore various facets of the Shannon entropy model**Understand** from its fundamental concepts to its applications in data analysis, information theory, and machine learning	CO 3
4	In the context of decision trees, how can a high level of information gain on training data lead to overfitting? What strategies can be employed to address this issue?	Understand	The learner will try to **recall**delve into various aspects of information gain, **Understand** from its fundamental concepts to its applications in decision trees and machine learning.	CO 3

5	Information gain is often applied to categorical features. How can it be extended to handle continuous or numerical features in decision tree algorithms?Provide an example of a dataset with both categorical and continuous features and discuss how information gain can guide feature selection.	Apply	The learner will try to **recall**delve into various aspects of information gain **Understand** from its fundamental concepts to **Apply** its applications in decision trees and machine learning.	CO 3
6	Provide a real-world application where information gain has been successfully used for feature selection and data splitting in a decision tree. What were the key considerations in applying this technique?	Apply	The learner will try to **recall** They encourage deeper **understand** and thoughtful consideration of how information gain **Apply** influences the construction of decision tree models.	CO 3
7	Are there alternative splitting criteria or measures that can be used instead of information gain in decision tree algorithms? What are the advantages and disadvantages of these alternatives?	Apply	The learner will try to **recall** They encourage deeper **Understand** and thoughtful consideration of how information gain **Apply** influences the construction of decision tree models.	CO 3
8	Provide examples of real-world applications where similarity-based learning techniques have been successfully applied. What unique challenges and opportunities do these applications present?	Understand	The learner will try to **recall** regarding the nuances **Understand** implications of using Count Vectorization in various NLP tasks	CO 3
9	Consider the interpretability and explainability of similarity-based learning models. How can you make the results and decisions of these models more transparent to users or stakeholders?	Analyze	The learner will try to **recall** They encourage deeper **understand** and thoughtful application of similarity-based learning **Apply** and **Analyze** techniques in various contexts.	CO 3

10	In sensitive or regulated domains, what ethical considerations should be taken into account when deploying similarity-based learning models that rely on potentially sensitive data?	Apply	The learner will try to **recall** They encourage deeper **understand** and thoughtful application of similarity-based learning techniques **Apply** in various contexts..	CO 3
PART-B LONG ANSWER QUESTIONS				
1	Explain the basic concept of a decision tree in machine learning. Describe how decision trees are constructed, and provide a step-by-step example using a real dataset. Discuss the advantages and disadvantages of decision trees compared to other machine learning algorithms. Analyze the key factors to consider when choosing a decision tree as a modeling tool in a given problem.	Understand	The learner will try to **recall** allowing students to delve deep into the theory, practical applications, and nuances **Understand** of this machine learning and decision-making tool.	CO 3
2	Compare and contrast different tokenization strategies, such as word-based, subword-based (e.g., Byte-Pair Encoding), and character-based tokenization. Discuss the advantages and limitations of each approach and provide examples of scenarios where specific strategies might be more suitable.	Understand	The learner will try to **recall**allowing students to delve deep into the theory, practical applications, and nuances **Undertsand** of this machine learning and decision-making tool.	CO 3

3	Explain the concept of overfitting in the context of decision trees. Describe how pruning techniques are used to mitigate overfitting. Provide detailed examples of overfitting and pruning in decision trees, including the impact of different pruning parameters. Discuss the trade-offs between a highly pruned tree and an unpruned tree.	Understand	The learner will try to **recall**allowing students to delve deep into the theory, practical applications, and nuances **Undertsand** of this machine learning and decision-making tool.	CO 3
4	Analyze methods for interpreting decision trees and extracting insights from them. Explain how to interpret the paths and nodes of a decision tree to make predictions and understand feature importance. Discuss visual representations, such as tree diagrams and variable importance plots, to aid in model interpretation. Provide examples of interpreting decision trees in real-world applications.	Analyze	The learner will try to **recall** these strategies, educators can help students develop a strong **understand** of how to interpret decision trees, extract insights,**Apply** and **Analyze** visualize their findings, preparing them for real-world data analysis tasks and model interpretation.	CO 3
5	Explain the concept of Shannon Entropy and its significance in information theory. Describe the mathematical formula for calculating Shannon Entropy for a discrete random variable. Provide real-world examples where Shannon Entropy can be applied to measure uncertainty or information content.	Understand	The learner will try to **recall** Students can explore these questions to gain a deep **understand** of how Shannon Entropy is used in information theory, machine learning, and diverse fields of study.	CO 3

6	Discuss how Shannon Entropy values relate to the predictability or randomness of a probability distribution. Provide detailed examples of probability distributions with low and high entropy values. Explain how the entropy of a uniform distribution compares to the entropy of a skewed distribution.	Understand	The learner will try to **recall**Students can explore these questions to gain a deep **understand** of how Shannon Entropy is used in information theory, machine learning, and diverse fields of study.	CO 3
7	Define cross-entropy and Kullback-Leibler (KL) divergence as measures of the difference between two probability distributions. Explain how cross-entropy can be used in machine learning for model evaluation, especially in classification tasks. Provide a mathematical derivation and practical examples of KL divergence calculations.	Understand	The learner will try to **recall**Students can explore these questions to gain a deep **understand** of how Shannon Entropy is used in information theory, machine learning, and diverse fields of study.	CO 3
8	Discuss the limitations and assumptions of Shannon Entropy as a model for measuring uncertainty. Explain situations where Shannon Entropy may not be an appropriate measure, such as when dealing with continuous variables or dependencies that it does not capture. Propose alternative measures or approaches to address these limitations.	Understand	The learner will try to **recall**Students can explore these questions to gain a deep **understand** of how Shannon Entropy is used in information theory, machine learning, and diverse fields of study.	CO 3

9	Explain the basic concept of the ID3 algorithm for decision tree construction. Discuss its objectives and significance in machine learning and data analysis. Provide a high-level overview of the key steps involved in ID3.	Understand	The learner will try to **recall** these strategies, educators can enhance students' understanding of the ID3 algorithm,**Understand** its objectives, and its significance in machine learning and data analysis.	CO 3
10	Describe the concepts of entropy and information gain, which are fundamental to the ID3 algorithm. Explain how entropy measures the impurity or disorder of a dataset and how information gain quantifies the reduction in entropy achieved by splitting data on a particular attribute. Provide mathematical formulas and intuitive explanations for these concepts.	Understand	The learner will try to **recall**to gain a comprehensive **understand** of how ID3 operates in decision tree learning.	CO 3
11	Discuss the different attribute selection criteria used by ID3, such as entropy, information gain, and gain ratio. Compare and contrast these criteria, highlighting their strengths and weaknesses. Provide scenarios where each criterion is preferred.	Understand	The learner will try to **recall**to gain a comprehensive **understand** of how ID3 operates in decision tree learning.	CO 3

12	Explain how the ID3 algorithm handles missing data and continuous attributes. Discuss strategies for imputing missing values and discretizing continuous attributes to make them suitable for decision tree construction. Provide examples to illustrate these processes.	Understand	The learner will try to **recall**to gain a comprehensive **understand** of how ID3 operates in decision tree learning	CO 3
13	Discuss the concept of pruning in decision tree construction and how it relates to ID3. Explain why overfitting is a concern in decision trees and describe techniques for reducing overfitting, such as reduced error pruning. Provide examples and trade-offs associated with pruning decisions.	Understand	The learner will try to **recall** to gain a comprehensive **understand** how ID3 operates in decision tree learning.	CO 3
14	Describe how ID3 handles categorical and nominal data, including the encoding of categorical attributes and the calculation of information gain for such attributes. Provide practical examples involving categorical data.	Understand	The learner will try to **recall**to gain a comprehensive **understand** of how ID3 operates in decision tree learning.	CO 3

15	Compare the ID3 algorithm with other decision tree algorithms, such as C4.5, CART, and Random Forests. Discuss the differences in their approaches, splitting criteria, and handling of various data types. Explain scenarios where ID3 may be preferred or less suitable compared to other algorithms.	Understand	The learner will try to **recall** to gain a comprehensive **understand** of how ID3 operates in decision tree learning.	CO 3
16	Explain the fundamental concept of similarity-based learning and its significance in machine learning. Describe the underlying idea of making predictions or decisions based on the similarity between data points. Provide examples of real-world applications where similarity-based learning is relevant.	Understand	The learner will try to **recall** to develop a comprehensive **understand** of how similarity-based learning is applied in various domains of machine learning.	CO 3
17	Discuss different similarity metrics commonly used in similarity-based learning, such as Euclidean distance, cosine similarity, Jaccard similarity, and Pearson correlation coefficient. Explain the mathematical formulations of these metrics and provide scenarios where each is appropriate.	Understand	The learner will try to **recall** to develop a comprehensive **understand** of how similarity-based learning is applied in various domains of machine learning.	CO 3

18	Discuss how similarity-based learning is employed in clustering algorithms, such as k-means and hierarchical clustering. Explain how data points are grouped together based on their similarity. Provide practical examples of clustering tasks in different domains and discuss the choice of similarity metric.	Apply	The learner will try to **recall** to develop a comprehensive **understand** of how similarity-based learning is **Apply** in various domains of machine learning.	CO 3
19	Explain the k-Nearest Neighbors (k-NN) algorithm and how it uses similarity measures to classify data points. Describe the impact of the choice of the value of k on the algorithm's performance. Provide examples of binary and multiclass classification problems that can be solved using k-NN and discuss the considerations when selecting a similarity metric?	Understand	The learner will try to **recall** to develop a comprehensive **understand** of how similarity-based learning is applied in various domains of machine learning.	CO 3
20	Explain the core concept of the Nearest Neighbor Algorithm and its significance in machine learning. Discuss the idea of using the similarity between data points for classification or regression. Provide examples of real-world applications where the Nearest Neighbor Algorithm is commonly used.	Apply	The learner will try to **recall** to develop a comprehensive **understand** of how the Nearest Neighbor Algorithm is **Apply** in various domains of machine learning.	CO 3

PART-C SHORT ANSWER QUESTIONS				
1	Explain the concept of a node in a decision tree.	Understand	The learner will try to **recall** the concept of a "node" in a decision tree is a fundamental element that plays a crucial role **understand** in representing and making decisions based on data.	CO 3
2	In decision trees, what is the purpose of a split or branching point?	Remember	—	CO 3
3	Describe the concept of information gain in decision trees.	Understand	The learner will try to **recall** the concept of "information gain" is a crucial aspect of **Understand** decision trees and is used to determine the best feature to split the data at each internal node.	CO 3
4	Explain the term "overfitting" in the context of decision trees.	Understand	The learner will try to **recall** when a decision tree model becomes excessively complex and **Understand** captures noise or random fluctuations in the training data rather than the underlying true patterns or relationships.	CO 3
5	How does the concept of entropy relate to decision tree construction?	Remember	—	CO 3
6	What is the purpose of feature importance in decision trees?	Remember	—	CO 3
7	How can decision trees be used for regression tasks, and what is the difference compared to classification tasks?	Understand	The learner will try to **recall** Decision trees can be used for both classification and **Understand** regression tasks	CO 3
8	How does Shannon Entropy quantify the amount of information in a random variable?	Remember	—	CO 3

9	Explain the concept of entropy in terms of uncertainty or disorder in a probability distribution.	Understand	The learner will try to **recall** Entropy, in the context of probability and **Understand** information theory, is a measure of uncertainty or disorder in a probability distribution	CO 3
10	What is the relationship between the number of possible outcomes and the entropy of a random variable?	Understand	The learner will try to **recall** Entropy, in the context of probability and **Understand** information theory, is a measure of uncertainty or disorder in a probability distribution	CO 3
11	Explain the concept of mutual information in the context of Shannon Entropy.	Understand	The learner will try to **recall** Mutual information is a concept in information theory that quantifies **Understand** the amount of information shared between two random variables.	CO 3
12	How is Shannon Entropy applied in information theory to measure the efficiency of data compression?	Remember	—	CO 3
13	What are some practical applications of Shannon Entropy in fields other than information theory?	Remember	—	CO 3
14	How is Shannon Entropy used in natural language processing to assess the informativeness of words in a text corpus?	Understand	The learner will try to **recall** this concept helps identify which words are more informative and **Understand** potentially carry more meaning in a given text corpus.	CO 3
15	What are the units of measurement for Shannon Entropy, and why are they relevant?	Remember	—	CO 3

16	Explain how Shannon Entropy can be used to calculate information gain in decision tree algorithms.	Understand	The learner will try to **recall** Shannon Entropy, often referred to **Understand** simply as entropy, is a concept from information theory that measures the uncertainty or impurity in a dataset or set of values.	CO 3
17	What is the main limitation of Shannon Entropy when dealing with continuous random variables, and how can it be addressed?	Remember	—	CO 3
18	How can Shannon Entropy be applied in data science to select relevant features for machine learning models?	Remember	—	CO 3
19	Explain the concept of a "splitting attribute" in the context of the ID3 algorithm.	Understand	The learner will try to **recall** the ID3 algorithm is a popular machine learning algorithm used for **Understand** building decision trees, primarily for classification tasks.	CO 3
20	Explain how ID3 handles missing values when constructing decision trees.	Understand	The learner will try to **recall** the ID3 algorithm is a popular machine learning algorithm used for **Understand** building decision trees, primarily for classification tasks.	CO 3

MODULE IV				
PROBABILITY BASED LEARNING				
PART A- PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS				
1	A certain medical test for a rare disease is 95% accurate in detecting true positives and 90% accurate in correctly identifying negatives. If the prevalence of the disease is 0.1% in the population, calculate the probability that a positive test result indicates an actual infection using Bayes' Theorem.	Evaluate	The learner will try to **recall** the application of Bayes' Theorem**Understand** to solve practical problems, make informed decisions, **Apply** and **Analyze** the impact of prior knowledge and **Evaluate** test accuracy on probability calculations.–	CO 4
2	In a decision-making context, how can Bayes' Theorem help you make optimal decisions by considering probabilities, costs, and benefits? Provide an example where Bayesian decision-making is applicable.	Understand	The learner will try to **recall** the application of Bayes' Theorem to solve practical problems, make informed decisions, and **understand** the impact of prior knowledge and test accuracy on probability calculations.	CO 4
3	As an investor, you are interested in predicting whether the stock price of a particular company will increase or decrease tomorrow. How can you use Bayesian prediction methods to factor in historical stock prices, market trends, and relevant news articles to make informed decisions?.	Understand	The learner will try to **recall** to think critically the consideration of relevant data, probabilistic reasoning, and **Understand** the incorporation of prior knowledge to make informed predictions and decisions..	CO 4

4	A telecommunications company wants to predict which of its customers are likely to churn (cancel their service) in the next month. How can Bayesian prediction techniques be employed to analyze customer behavior data, such as call history and customer complaints, to identify potential churners?	Understand	The learner will try to **recall** to think critically the consideration of relevant data, probabilistic reasoning, and **Understand** the incorporation of prior knowledge to make informed predictions and decisions..	CO 4
5	You are developing a spam email filter. Explain how Bayesian prediction, specifically using the Naive Bayes classifier, can be used to classify incoming emails as spam or not. What are the key features and probabilities involved in this prediction?	Apply	The learner will try to **recall**to think critically the consideration of relevant data, probabilistic reasoning, and **understand** the incorporation of prior knowledge to **Apply** make informed predictions and decisions.	CO 4
6	A manufacturing company is concerned about defects in its products. How can Bayesian prediction be used to assess the likelihood of a product being defective based on historical quality control data and information about the manufacturing process?	Understand	The learner will try to **recall**to think critically the consideration of relevant data, probabilistic reasoning, and **understand** the incorporation of prior knowledge to make informed predictions and decisions.	CO 4
7	City planners want to predict traffic congestion in different areas of a city throughout the day. Explain how Bayesian prediction models can incorporate traffic sensor data, historical traffic patterns, and real-time events (e.g., accidents, road closures) to make traffic flow predictions.	Analyze	The learner will try to **recall** to think critically the consideration of relevant data **Understand** probabilistic reasoning, **Apply** and **Analyze** the incorporation of prior knowledge to make informed predictions and decisions.	CO 4

8	Imagine you have a continuous random variable representing the height of individuals in a population. How would you use a Probability Density Function (PDF) to estimate the probability that a randomly selected person is taller than a certain height? What information would you need to create such a PDF?	Analyze	The learner will try to **recall** by **understand** requiring individuals to **Apply** knowledge of PDFs in various contexts, from **Analyze** probability calculations to data analysis and decision-making in real-world scenarios.	CO 4
9	You have two continuous random variables, X and Y, and their joint PDF. How can you calculate the conditional probability P(X ¿ a — Y ¿ b), where 'a' and 'b' are specific values? What does this conditional probability tell you about the relationship between X and Y?.	Understand	The learner will try to **recall** by requiring individuals to apply knowledge of PDFs in various contexts **Understand** from probability calculations to data analysis and decision-making in real-world scenarios.	CO 4
10	In a manufacturing process, you have a measurement that follows a continuous distribution. You want to set quality control limits based on a desired level of defectiveness. How can you use the PDF to determine these control limits and minimize defects?	Understand	The learner will try to **recall** by requiring individuals to apply knowledge of PDFs in various contexts **understand** from probability calculations to data analysis and decision-making in real-world scenarios.	CO 4

PART-B LONG ANSWER QUESTIONS				
1	Explain the fundamental concept of Bayes' Theorem. Provide the mathematical formula and discuss the terms involved, such as prior probability, likelihood, marginal likelihood, and posterior probability. How does Bayes' Theorem help in updating beliefs or probabilities?	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the fundamental concept of Bayes' Theorem, its mathematical formulation, and its practical utility in updating beliefs or probabilities in various domains.	CO4
2	Describe a medical scenario where Bayes' Theorem can be applied effectively for diagnosis. Include details about the prior probability of a disease, the sensitivity and specificity of a diagnostic test, and how Bayes' Theorem can help calculate the probability of having the disease given a positive test result.	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the fundamental concept of Bayes' Theorem, its mathematical formulation, and its practical utility in updating beliefs or probabilities in various domains	CO 4
3	Discuss how Bayes' Theorem is used in spam email filtering, specifically with the Naive Bayes classifier. Explain the concept of feature selection, likelihood estimation, and prior probabilities in the context of classifying emails as spam or not.	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the fundamental concept of Bayes' Theorem, its mathematical formulation, and its practical utility in updating beliefs or probabilities in various domains	CO 4
4	Explain how the Naive Bayes classifier works in machine learning. Describe the "naive" assumption it makes about feature independence. Provide an example of text classification, such as sentiment analysis or spam detection, where Naive Bayes is applied.	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the fundamental concept of Bayes' Theorem, its mathematical formulation, and its practical utility in updating beliefs or probabilities in various domains	CO 4

5	Compare and contrast the Bayesian and frequentist approaches to probability and statistical inference. Discuss the philosophical and practical differences between the two approaches, highlighting situations where one might be more suitable than the other.	Analyze	The learner will try to **recall** Bayes' Theorem **Understand** its applications, allowing students to explore the theory, practical uses, **Apply**and **Analyze** implications of Bayesian inference in various contexts.	CO 4
6	Explain the concept of conditional independence in probability theory. How does it differ from unconditional independence? Provide an example that illustrates the difference between these two types of independence	Understand	The learner will try to **recall** the concept of Principle component Analysis and **Demonstrate** its application	CO 4
7	Describe how conditional independence is used in the context of Bayesian networks. Explain the concept of a Bayesian network and how it represents probabilistic dependencies among random variables. Provide an example of a simple Bayesian network and explain how conditional independence assumptions are made within it.	Understand	The learner will try to **recall** the concept AGNES Algorithm and **Demonstrate** its application	CO 4
8	Discuss the idea of factorization in the context of probability distributions. How can a joint probability distribution be factorized into conditional probabilities and marginal probabilities? Provide a mathematical representation of factorization for a simple probability distribution.	Understand	The learner will try to **recall** these strategies, educators can help students**understand** the idea of factorization in probability distributions and its significance in simplifying complex probabilistic models.	CO 4

9	Explain the Factorization Theorem for Bayesian networks. How does it relate to conditional independence in the context of graphical models? Describe the role of the network structure and the conditional probability tables (CPTs) in the factorization of a joint distribution.	Understand	The learner will try to **recall** these strategies, educators can help students**understand** the idea of factorization in probability distributions and its significance in simplifying complex probabilistic models.	CO 4
10	Describe the process of probabilistic inference in a Bayesian network. How can conditional independence assumptions and factorization be leveraged to efficiently compute probabilities of interest in a large graphical model? Explain the role of inference algorithms like variable elimination.	Understand	The learner will try to **recall** conditional independence and factorization in probability theory and graphical models, allowing students to delve into the theory and practical applications of these concepts.	CO 4
11	Explain the fundamental concept of the Naïve Bayes model in machine learning. Describe how it uses Bayes' Theorem and the assumption of feature independence. Provide examples of real-world applications where the Naïve Bayes model is commonly used.	Understand	The learner will try to **recall** the concept of K-Mode Clustering and **Demonstrate** it on the given sample	CO 4
12	Explain the various types of Naïve Bayes classifiers, such as Gaussian Naïve Bayes, Multinomial Naïve Bayes, and Bernoulli Naïve Bayes. Discuss when each type is most suitable and provide practical examples.	Understand	The learner will try to **recall** the concept of K-Mode Clustering and **Demonstrate** it on the given sample	CO 4

13	Discuss the Bayesian probability framework and how it forms the foundation for the Naïve Bayes model. Explain the role of prior probabilities and likelihood in probability estimation. How do these probabilities contribute to the final prediction?	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the Bayesian probability framework, the roles of prior probabilities and likelihood, and how they contribute to making informed predictions and decisions in various domains.	CO 4
14	Describe the problem of zero probabilities in the Naïve Bayes model and how Laplace smoothing addresses this issue. Provide the mathematical formula for Laplace smoothing and discuss its impact on probability estimation.	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the Bayesian probability framework, the roles of prior probabilities and likelihood, and how they contribute to making informed predictions and decisions in various domains	CO 4
15	Discuss the importance of feature selection and feature engineering in Naïve Bayes modeling. How can you preprocess and select relevant features to improve model performance? Provide examples of feature engineering techniques.	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the Bayesian probability framework, the roles of prior probabilities and likelihood, and how they contribute to making informed predictions and decisions in various domains	CO 4
16	Explain in detail how the Naïve Bayes model is applied to the problem of spam email filtering. Discuss the key features used, the training process, and the challenges associated with adapting to evolving spamming techniques.	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the Bayesian probability framework, the roles of prior probabilities and likelihood, and how they contribute to making informed predictions and decisions in various domains	CO 4

17	Explain how to determine the probability distribution of a transformed continuous random variable (e.g., Y = 2X + 3) given the PDF of the original random variable X. Use a mathematical example to illustrate this transformation and its effect on the PDF.	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the Bayesian probability framework, the roles of prior probabilities and likelihood, and how they contribute to making informed predictions and decisions in various domains	CO 4
18	Describe the relationship between a PDF and the Cumulative Distribution Function (CDF) of a continuous random variable. How can the CDF be derived from the PDF, and how are they used together to compute probabilities?	Understand	The learner will try to **recall**Probability Density Functions, allowing students to explore the theory, applications, and practical considerations associated with continuous probability distributions.	CO 4
19	Discuss how to compute the expectation (mean) and variance of a continuous random variable using its PDF. Include the relevant mathematical formulas and explain their significance in characterizing the distribution.	Understand	The learner will try to **recall**Probability Density Functions, allowing students to explore the theory, applications, and practical considerations associated with continuous probability distributions.	CO 4
20	Describe methods for approximating a PDF from a dataset of continuous observations. Explain the use of histograms, kernel density estimation, and parametric PDF fitting. Discuss the advantages and limitations of each approach.	Understand	The learner will try to **recall** Probability Density Functions, allowing students **understand** to explore the theory, applications, and practical considerations associated with continuous probability distributions.	CO 4
PART-C SHORT ANSWER QUESTIONS				

1	Explain the key components of Bayes' Theorem, including prior probability, likelihood, marginal likelihood, and posterior probability.	Understand	The learner will try to **recall** Bayes' Theorem is a fundamental concept in probability theory and **Understand** statistics that describes how to update our beliefs or probabilities based on new evidence or information	CO 4
2	What is Bayes' Theorem, and what is its primary purpose in probability theory?	Understand	The learner will try to **recall** Bayes' Theorem is a fundamental concept in probability theory and **Understand** statistics that describes how to update our beliefs or probabilities based on new evidence or information.	CO 4
3	How does Bayes' Theorem provide a framework for updating beliefs or probabilities based on new evidence or observations?	Remember	—	CO 4
4	Define likelihood in the context of Bayes' Theorem. How does it describe the probability of observing data given a particular hypothesis?	Remember	—	CO 4
5	What does the term "posterior probability" refer to in Bayes' Theorem, and how is it calculated based on the prior probability and likelihood?	Remember	—	CO 4
6	Describe a real-world situation where Bayes' Theorem can be applied to make informed decisions or predictions.	Remember	—	CO 4
7	What is conditional independence in probability theory, and how does it differ from unconditional independence?	Remember	—	CO 4

8	Define the Markov blanket of a node in a Bayesian network and its significance in representing conditional independence.	Remember	—	CO 4
9	What is factorization in the context of probability distributions, and why is it an important concept?	Remember	—	CO 4
10	Explain the concept of D-Separation in Bayesian networks and how it determines conditional independence relationships.	Understand	The learner will try to **recall** D-Separation is a crucial tool for **understand** probabilistic dependencies and making probabilistic inferences in Bayesian networks.	CO 4
11	What is Laplace smoothing (additive smoothing), and how does it address the issue of zero probabilities in the Naïve Bayes model?.	Remember	—	CO 4
12	Explain why the Naïve Bayes classifier makes a "naïve" assumption about feature independence and how this assumption simplifies the model.	Understand	The learner will try to **recall** this assumption simplifies the model significantly and allows for efficient and **Understand** straightforward probabilistic calculations.	CO 4
13	What are some common extensions or variations of decision tree algorithms, and how do they address limitations of basic decision trees?	Remember	—	CO 4
14	Can you name a variation of the Naïve Bayes classifier, and Explain how it differs from the traditional Naïve Bayes model?	Understand	The learner will try to **recall** in terms of the type of data it's designed to handle and **Understand** the underlying probability distribution it assumes.	CO 4

15	Name a variation of Bayesian networks known as dynamic Bayesian networks (DBNs) and explain their utility in modeling temporal dependencies.	Understand	The learner will try to **recall** DBNs are particularly useful when dealing with sequential data, time series, or **Understand** any data where the order of events or observations matters.	CO 4
16	What does the Probability Density Function (PDF) represent in probability theory and statistics?	Remember	—	CO 4
17	Explain the primary difference between the PDF of a continuous random variable and the probability mass function (PMF) of a discrete random variable.	Understand	The learner will try to **recall** the primary difference between the Probability Density Function (PDF) of a continuous random variable and **Understand** the Probability Mass Function (PMF) of a discrete random variable	CO 4
18	What are the key properties that a valid PDF must satisfy?	Remember	—	CO 4
19	How is the probability of a continuous random variable falling within a specific range calculated using its PDF?	Remember	—	CO 4
20	What does it imply if two continuous random variables have the same PDF?	Remember	—	CO 4
MODULE V				
ERROR BASED LEARNING				
PART A-PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS)				
1	In a supervised learning scenario, you have a classification model that frequently misclassifies a specific class. How would you analyze the misclassified instances to understand the sources of error?	Analyze	The learner will try to **recall** encourage a deeper **Understand** of simple linear regression **Apply** its appropriate use, assumptions, evaluation, and **Analyze** real-world applications.	CO 5

2	Select a specific industry or domain (e.g., healthcare, marketing, finance) and discuss how multivariable linear regression with gradient descent is commonly used to address practical problems in that domain.	Understand	The learner will try to **recall**to encourage a deeper exploration of **understand** multivariable linear regression with gradient descent, its challenges, optimization strategies, and practical applications across different domains..	CO 5
3	You have a reasonably performing multivariable regression model. What strategies or techniques would you explore to further improve its predictive accuracy? Can you share an example from your experience where a small improvement in model performance had a significant impact on decision-making or outcomes?	Understand	The learner will try to **recall**to encourage a deeper exploration of **Understand** multivariable linear regression with gradient descent, its challenges, optimization strategies, and practical applications across different domains.	CO 5
4	In a binary classification problem, explain the difference between accuracy, precision, recall, and F1-score as error metrics. When might you prioritize one metric over the others?Can you provide an example where accuracy may not be an appropriate measure of model performance, and another metric is more informative?.	Apply	The learner will try to **recall** various aspects of measuring error**Understand** from selecting appropriate metrics to handling uncertainty and **Apply** ethical considerations in error analysis.	CO 5
5	When comparing multiple models, what considerations should guide your choice of error metrics?Provide an example where different error metrics led to different model preferences, and discuss the implications.	Analyze	The learner will try to **recall**to think critically about exploring various aspects of measuring error	CO 5

6	Provide examples of real-world applications where knowledge of error surfaces and optimization is critical, such as in deep learning or engineering design.How does the knowledge of error surfaces impact decision-making in these applications?	Apply	The learner will try to **recall**to think critically **Understand** by exploring the concepts of error surfaces, their optimization, and their **Apply** practical implications in various domains.	CO 5
7	Are there emerging technologies or methodologies that are changing the way error surfaces are analyzed and optimized in machine learning and optimization? How might advancements in optimization algorithms and computational resources influence the handling of complex error surfaces?	Analyze	The learner will try to **recall** to think critically **Understand** by exploring the concepts of error surfaces, their optimization**Apply** and **Analyze** their practical implications in various domains.	CO 5
8	Are there emerging technologies, methodologies, or tools that are changing the way error surfaces are analyzed and optimized in machine learning and optimization? How might advancements in optimization algorithms and computational resources impact the handling of complex error surfaces in the future?	Apply	The learner will try to **recall** to think critically **understand** by exploring the concepts of error surfaces, their role in optimization, and their practical implications **Apply** in various domains.	CO 5

9	In real-time systems (e.g., autonomous vehicles, natural language processing), how can error-based learning techniques be applied to adapt to changing conditions and improve decision-making? What are the challenges associated with implementing error-based learning in real-time systems, and how can they be mitigated?	Understand	The learner will try to **recall** to think critically skills **Understand** by exploring the principles, applications, challenges, and ethical considerations related to error-based learning in various contexts.	CO 5
10	In human-in-the-loop learning systems, how can errors and feedback from human operators be leveraged to enhance machine learning models? What roles do humans play in identifying errors, providing feedback, and guiding the learning process?	Analyze	The learner will try to **recall** to think critically skills **Understand** by exploring the principles, applications, challenges, **Apply**and **Analyze** ethical considerations related to error-based learning in various contexts.	CO 5

PART-B LONG ANSWER QUESTIONS				
1	Discuss the Importance of Assumptions in Simple Linear Regression and Their Implications for Interpretation.	Understand	The learner will try to **recall** these strategies, educators can help students appreciate **Understand** the importance of assumptions in simple linear regression and equip them with the skills to interpret regression results in a meaningful and contextually relevant manner.	CO 5
2	Describe the Process of Hypothesis Testing in Simple Linear Regression, Including the Role of p-Values and Confidence Intervals.	Understand	The learner will try to **recall** these strategies, educators can help students appreciate **Understand** the importance of assumptions in simple linear regression and equip them with the skills to interpret regression results in a meaningful and contextually relevant manner.	CO 5
3	Explain the Concept of Error Measurement in Statistics and Its Significance in Data Analysis.	Understand	The learner will try to **recall** to explore Bidirectional Recurrent Neural Networks (RNNs) **understand** from various angles, including architecture, training, applications, challenges, and future directions.	CO 5
4	Describe the Types of Error Measurement Metrics Used in Regression Analysis, Such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE)..	Understand	The learner will try to **recall** to explore Bidirectional Recurrent Neural Networks (RNNs) **understand** from various angles, including architecture, training, applications, challenges, and future directions	CO 5

5	Explain the Concept of Bias and Variance in Error Measurement and Their Trade-off in Model Selection?	Understand	The learner will try to **recall** to explore Bidirectional Recurrent Neural Networks (RNNs) **understand** from various angles, including architecture, training, applications, challenges, and future directions	CO 5
6	Explain the Concept of an Error Surface in Machine Learning. What Role Does It Play in Model Training?	Understand	The learner will try to **recall** to explore Bidirectional Recurrent Neural Networks (RNNs) **understand** from various angles, including architecture, training, applications, challenges, and future directions	CO 5
7	How Does the Shape of the Error Surface Impact the Performance and Behavior of Optimization Algorithms? Provide Examples of Different Error Surface Shapes?	Apply	The learner will try to **recall** to explore Bidirectional Recurrent Neural Networks (RNNs) **understand** from various angles, including architecture, training, applications, challenges, and **Apply** future directions	CO 5
8	Discuss the Challenges and Strategies for Dealing with High-Dimensional Error Surfaces in Machine Learning.	Understand	The learner will try to **recall** to explore Bidirectional Recurrent Neural Networks (RNNs) **understand** from various angles, including architecture, training, applications, challenges, and future directions	CO 5
9	How Does the Shape of the Error Surface Impact the Performance and Behavior of Optimization Algorithms? Provide Examples of Different Error Surface Shapes.	Understand	The learner will try to **recall** to explore Bidirectional Recurrent Neural Networks (RNNs) **understand** from various angles, including architecture, training, applications, challenges, and future directions	CO 5

10	Explain the Concept of Multivariable Linear Regression and How It Differs from Simple Linear Regression. Then, Describe the Role of Gradient Descent in Multivariable Linear Regression.	Understand	The learner will try to **recall** these strategies, educators can help students **understand** the concept of multivariable linear regression, its differences from simple linear regression, and the role of gradient descent in optimizing regression models with multiple predictors.	CO 5
11	Illustrate the Steps Involved in Implementing Multivariable Linear Regression with Gradient Descent. Include Initialization, the Gradient Descent Iteration, and the Stopping Criteria.	Understand	The learner will try to **recall** these strategies, educators can help students through the steps of implementing multivariable linear regression with gradient descent and providing practical examples, educators can enhance their **Understand** of this essential machine learning technique	CO 5
12	Discuss the Challenges and Considerations in Multivariable Linear Regression with Gradient Descent, Particularly with Regard to Overfitting, Learning Rate, and Feature Scaling.	Understand	The learner will try to **recall** these strategies, educators can help students through the steps of implementing multivariable linear regression with gradient descent and providing practical examples, educators can enhance their **Understand** of this essential machine learning technique	CO 5

13	Explain the Concept of Regularization in Machine Learning and Discuss Two Common Techniques: L1 (Lasso) and L2 (Ridge) Regularization. How Do They Differ, and When Would You Use Each?	Understand	The learner will try to **recall** these strategies, educators can help students grasp the concept of regularization, **understand** the differences between L1 and L2 regularization, and learn when to apply each technique effectively in machine learning projects.	CO 5
14	Describe the Concept of Ensemble Learning in Machine Learning. Explain Bagging and Boosting, Providing Examples of Algorithms for Each Approach. How Do They Improve Model Performance?	Understand	The learner will try to **recall** these strategies, educators can help students grasp the concept of regularization, **understand** the differences between L1 and L2 regularization, and learn when to apply each technique effectively in machine learning projects.	CO 5
15	Discuss the Concept of Deep Learning and Its Relationship with Artificial Neural Networks (ANNs). Explain the Advantages and Limitations of Deep Learning, and Provide Examples of Real-World Applications.	Understand	The learner will try to **recall** these strategies, educators can help students grasp **understand** the concept of deep learning, its relationship with ANNs, and its real-world applications while understanding the advantages and limitations associated with this powerful machine learning paradigm.	CO 5
16	Define Error-Based Learning and Explain Its Significance in Machine Learning. How Does It Differ from Other Learning Paradigms?	Understand	The learner will try to **recall**these strategies, educators can help students **understand** the concept of error-based learning, its significance in machine learning, and how it differs from other learning paradigms based on objectives and approaches.	CO 5

17	Describe the Basic Components of Error-Based Learning. What Are the Input Data, Predictions, Errors, and Learning Algorithms, and How Do They Interact?	Understand	The learner will try to **recall**these strategies, educators can help students **understand** the concept of error-based learning, its significance in machine learning, and how it differs from other learning paradigms based on objectives and approaches.	CO 5
18	Explain the Role of Loss Functions in Error-Based Learning. What Are Some Common Loss Functions, and How Are They Chosen for Different Tasks?	Understand	The learner will try to **recall**these strategies, educators can help students **understand** the concept of error-based learning, its significance in machine learning, and how it differs from other learning paradigms based on objectives and approaches.	CO 5
19	Explain the Core Concept of Error-Based Learning in Machine Learning. How Does It Differ from Other Learning Paradigms, and What Is Its Significance in Model Training?	Understand	The learner will try to **recall**these strategies, educators can help students **understand** the concept of error-based learning, its significance in machine learning, and how it differs from other learning paradigms based on objectives and approaches.	CO 5
20	Discuss the Concept of Bias and Variance in Error-Based Learning. How Do These Concepts Affect Model Performance, and What Strategies Can Be Employed to Balance Them?	Understand	The learner will try to **recall**these strategies, educators can help students **understand** the concept of error-based learning, its significance in machine learning, and how it differs from other learning paradigms based on objectives and approaches.	CO 5

PART-C SHORT ANSWER QUESTIONS				
1	How does stochastic gradient descent (SGD) differ from standard gradient descent in error-based learning?	Remember	—	CO 5
2	What are some real-world applications of error-based learning?	Remember	—	CO 5
3	How do adaptive learning rate methods enhance error-based learning in optimization?	Understand	The learner will try to **recall** Adaptive learning rate methods are optimization techniques used in machine learning and **Understand** neural network training to enhance error-based learning.	CO 5
4	Why is error-based learning considered fundamental in modern machine learning?	Understand	The learner will try to **recall** Adaptive learning rate methods are optimization techniques used in machine learning and **Understand** neural network training to enhance error-based learning.	CO 5
5	What is simple linear regression?	Remember	—	CO 5
6	What is the purpose of simple linear regression?	Remember	—	CO 5
7	Explain the relationship in simple linear regression represented mathematically?	Understand	The learner will try to **recall** simple linear regression **Understand** mathematically represents the relationship between two variables using a linear equation.	CO 5
8	What is the role of the intercept in simple linear regression?	Remember	—	CO 5
9	How is the best-fit line determined in simple linear regression?	Remember	—	CO 5

10	In what situations is simple linear regression most suitable?	Understand	The learner will try to **recall** simple linear regression **Understand** mathematically represents the relationship between two variables using a linear equation.	CO 5
11	How is the Root Mean Square Error (RMSE) different from the Mean Absolute Error (MAE)?	Understand	The learner will try to **recall** Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) are both metrics used to **Understand** measure the accuracy of predictions or models.	CO 5
12	What does a lower MAE or RMSE value indicate about a predictive model's performance?	Understand	The learner will try to **recall** Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) are both metrics used to **Understand** measure the accuracy of predictions or models.	CO 5
13	How is the coefficient of determination used to assess error in regression models?	Remember	—	CO 5
14	In classification tasks, what is the purpose of measuring error using metrics like accuracy and F1 score?	Remember	—	CO 5
15	What is the concept of "bias" when measuring error in machine learning models?	Remember	—	CO 5
16	How can cross-validation be used to estimate error in machine learning models?	Remember	—	CO 5
17	Explain the cost function in multivariable linear regression, and how is it related to gradient descent?	Understand	The learner will try to **recall** When using gradient descent in multivariable linear regression **Understand** several common challenges or issues may arise.	CO 5

18	How does gradient descent work in the context of multivariable linear regression?	Remember	—	CO 5
19	Explain some common challenges or issues one may encounter when using gradient descent in multivariable linear regression?	Understand	The learner will try to **recall** When using gradient descent in multivariable linear regression **Understand** several common challenges or issues may arise	CO 5
20	How can one assess the quality of a multivariable linear regression model after training with gradient descent?	Remember	—	CO 5

Course Coordinator:
Ms. S.Viharika

HOD CSE(AI & ML)